Boosting Robot Learning and Control with Domain Constraints

Hang Yin[1,2], Francisco S. Melo[1], Aude Billard[2] and Ana Paiva[1]
[1]GAIPS, INESC-ID and Instituto Superior Técnico, Universidade de Lisboa
[2]Learning Algorithms and Systems Laboratory, École Polytechnique Fédérale de Lausanne

I. Introduction

In this short paper, we exploit robotics domain knowledge, be it acquired from humans or self-organization, to alleviate learning and control challenges from directly dealing with raw demonstrations or sparse reward signals. We take a unified latent variable perspective in incorporating domain constraints. The latent variables are regarded as task parameters or representations, which rationalize task observations with a generative model. The constraints can thus be specified with structured latent variables. Different from many related works, we explore latent structures that are computationally feasible and robotics-oriented to facilitate both task learning and control synthesis. The paper will briefly discuss adopted structures ranging from parameter dependency, modality and dynamical associativity, extending imitation learning such as inverse optimal control and deep generative models. The framework is shown to be effective in a range of manipulation tasks, including 1) learning variable impedance controllers in robotic handwriting; 2) boosting motion synthesis for writing novel symbols; 3) reasoning an internal model to score a ball-target under malfunctioning visual input.

II. Related Work

Embedding domain constraints has been explored in various forms. [13] embeds a locally-linear dynamical system for learning a pixel-based inverted pendulum task. In [5], a similar system improves the training error propagation, filtering latent variables correlated to the pendulum angular velocity. [1] proposes to parameterize tasks with representations from different reference frames, demonstrating improved generalization in novel task configurations. The equivalence between discrete Bellman iterations and convolution-pooling operations is employed in [11]. The model architecture achieves notably better adaptation in pixel-based navigation tasks. Besides crafting the model architecture, researchers also look into auxiliary constraints and objectives. [6] enforces parameter constraints inspired from the Lypunov criterion to learn reaching motion with assured stability. [4] optimizes local smoothness to extract the tangential space of a skill manifold for generalization based on geodesic distance. Recently, research efforts have also been made on acquiring domain priors from the data. In [10], human demonstrations are used to estimate time-invariant dynamical systems, which in turn initialize and shape the motion refinement. More formal treatments under a transfer learning framework also showcase successes by learning simulation-based tasks [14][12] or a flexible policy initialization [3].

In this paper, we focus on imposing structures to an explicit representation of latent variables. This is realized by designing factored parameter distribution and general auxiliary objectives with a minor additional training cost. Moreover, unlike the transfer learning works, priors are usually not encoded as target policies but intermediate models such as cost functions, feature and modality transformations.

III. Approach and Current Results

We assume there exists an abstract latent representation $\boldsymbol{z}$ which conditions the generation of $\boldsymbol{x}$ which denotes the observations of interests, e.g., motion or visual frames, formulating the likelihood $p(\boldsymbol{x})$ and its variational lower bound as:

$$\begin{aligned} p(\boldsymbol{x}) &= \int p(\boldsymbol{x}|\boldsymbol{z})p_0(\boldsymbol{z})d\boldsymbol{z} \\ \log p(\boldsymbol{x}) &\geq -\mathbb{E}_q(\boldsymbol{z}|\boldsymbol{x})(\log p(\boldsymbol{x}|\boldsymbol{z})) - \mathrm{KL}(q(\boldsymbol{z}|\boldsymbol{x})\|p_0(\boldsymbol{z})) \end{aligned} \tag{1}$$

where $p_0(\boldsymbol{z})$ and $q(\boldsymbol{z}|\boldsymbol{x})$ denote the prior and approximate posterior distributions. The generative model $p(\boldsymbol{x}|\boldsymbol{z})$ is instanced as an energy-based model $p(\boldsymbol{x}|\boldsymbol{z}) \propto e^{-\mathcal{J}(\boldsymbol{x},\boldsymbol{z})}$ [17].

Here $\boldsymbol{z}$ can be cast as the task parameter of $\mathcal{J}(\cdot)$. In light of $q(\boldsymbol{z}|\boldsymbol{x})$, another way is to view $\boldsymbol{z}$ as a latent feature which maps raw $\boldsymbol{x}$ to a low-dimensional space. We discuss imposed constraints on $\boldsymbol{z}$ and obtained results in robot applications.

A. Parameter Dependency

We consider a quadratic $\mathcal{J}(\boldsymbol{x}) = (\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Lambda}(\boldsymbol{x}-\boldsymbol{\mu})$ with $\boldsymbol{z} = \{\boldsymbol{\mu}, \boldsymbol{\Lambda}\}$. The objective in effect encodes motion trajectories $\boldsymbol{x}$ with a reference $\boldsymbol{\mu}$ and tracking desirability $\boldsymbol{\Lambda}$. Meanwhile, $\boldsymbol{\Lambda}$ can be regarded as a force control proxy because of the popular heuristics correlating trajectory variance and compliance design [2][8]. Here, we adopt a description of $\boldsymbol{\Lambda}$ which depends on $\boldsymbol{\mu}$, resulting a factored p_0:

$$p_0(\boldsymbol{z}) = p_0(\boldsymbol{\Lambda}|\boldsymbol{\mu})p_0(\boldsymbol{\mu}) \tag{2}$$

This can facilitate a sampling-based inverse optimal control to efficiently optimize the likelihood (1), avoiding dealing with ill-posed parameter constraints [15]. Moreover, the dependency implies representing orthogonal control components in a local frame along the trajectory. This is known to be an important and intuitive task space decomposition in classic robotics, such as hybrid force control [9]. In a robotic handwriting domain, a variable impedance controller is developed and the writing motion on a moving surface (Figure 1).

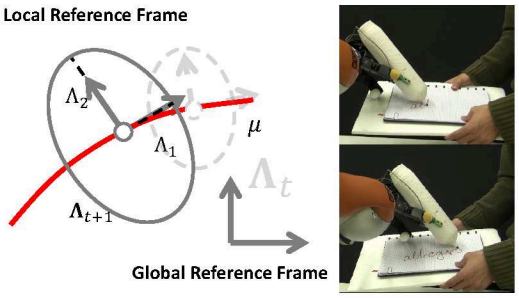

Fig. 1: Enforcing a dependency between task parameters to constrain the impedance representation: the arm-hand system performs compliant cursive handwriting, regulating the contact with a surface, whose orientation varies during the process.

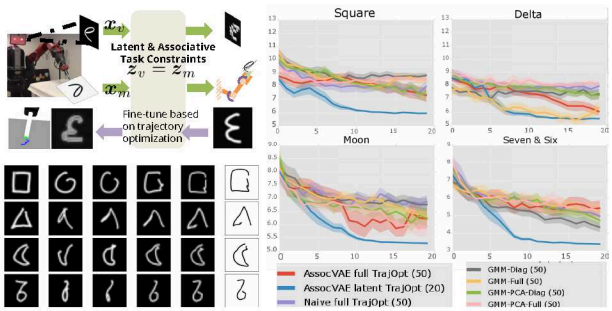

Fig. 2: Using learned associative constraints in warm-starting the trajectory optimization: the associative encoders (AssocVAE Full/Latent) provide an informed initialization and low-dimensional manifolds (AssocVAE Latent) to explore the arm joint motion for writing novel non-alphabetical symbols, outperforming the baselines based on GMMs with less sampling trajectories (numbers in parentheses after the legend names).

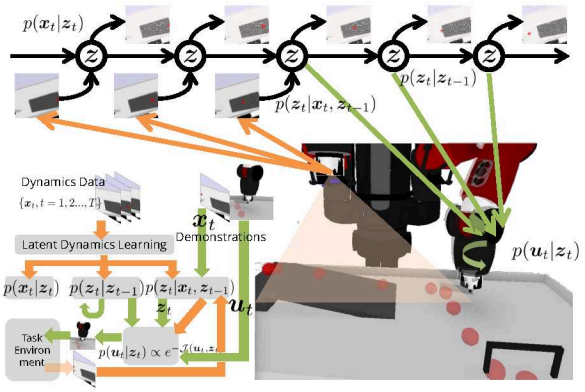

Fig. 3: Embedding a dynamical relation with z_t as an abstract representation of frames $x_{0:t}$ (upper). The retained feature transformation and dynamics are exploited to learn a model-based control in the latent space (lower left): after accessing initial frames, a robot arm anticipates the rolling ball movement and score a goal without visual input (lower right).

B. Identical Modal Representation

We consider task examples $\boldsymbol{x}$ of different modalities, e.g., joint trajectory $\boldsymbol{x}_m$ and camera pixels x_v that form handwriting letters. The redundant description is similar to observing the pose of an object from different perspectives. This inspires an identical representation of $\boldsymbol{z}$ to abstract both $\boldsymbol{x}_m$ and $\boldsymbol{x}_v$. We enforce this domain constraint by adding a symmetrical KL-divergence to the variational objective in (1):

$$\begin{aligned} & q_m(\boldsymbol{z}|\boldsymbol{x_m}) = q_v(\boldsymbol{z}|\boldsymbol{x_v}) \\ \Rightarrow & \text{Diff}(q_m, q_v) = \text{KL}(q_m \| q_v) + \text{KL}(q_v \| q_m) \end{aligned} \tag{3}$$

This yields associative variational auto-encoders [16] when we parameterize posterior q and generative models $p(\boldsymbol{x}_m|\boldsymbol{z})$ and $p(\boldsymbol{x}_v|\boldsymbol{z})$ with neural models as the ones in [7]. The association can be exploited in control synthesis:

$$\begin{aligned} \boldsymbol{x}_m^* &= \underset{x_m}{\text{argmin}} \, \|\text{Img}(\boldsymbol{x}_m) - \boldsymbol{x}_v\|^2 + \lambda \text{KL}[q(\boldsymbol{z}|\boldsymbol{x}_m) \| p_0] \\ \boldsymbol{x}_m^{\text{Init}} &\sim p(\boldsymbol{x}_m|\boldsymbol{x}_v) = \int p(\boldsymbol{x}_m|\boldsymbol{z}) q(\boldsymbol{z}|\boldsymbol{x}_v) d\boldsymbol{z} \end{aligned} \tag{4}$$

where the shared prior p_0 and posterior $q(\boldsymbol{z}|\boldsymbol{x}_m)$ provide informative initialization and bias in adapting motion $\boldsymbol{x}_m$ to match a novel letter image $\boldsymbol{x}_v$, with $\text{Img}(\cdot)$ and $\|\cdot\|$ denoting the obtained image and an Euclidean norm. The KL term encourages to search in alignment with the manifold and is weighted by λ. As is shown in Figure 2, these constraints help to establish a rapid and stable convergence with less exploring trajectory samples.

C. Latent Dynamics

We consider task examples that can be temporally factored $\boldsymbol{x} = \boldsymbol{x}_{0:T}$. This motivates to embed a latent dynamics constraint over $\boldsymbol{z} = \boldsymbol{z}_{0:T}$ if $\boldsymbol{z}_t$ is taken as a belief representation of the observation history $\boldsymbol{x}_{0:t}$. Following the case of bridging modalities in (3), we capture the temporal relation by matching its estimation of the posterior and prior latent dynamics:

$$q(\boldsymbol{z}_t|\boldsymbol{z}_{t-1}, \boldsymbol{x}_t) = p_0(\boldsymbol{z}_t|\boldsymbol{z}_{t-1}) \Rightarrow \text{KL}(q \| p_0) \tag{5}$$

Similar to the auto-encoder settings, high-dimensional patterns can be reconstructed by reasoning about the latent dynamics. The dynamics learning also enables a model-based control when sensory input is not available (Figure 3).

IV. Future Plan

We envision future research steps of utilizing domain constraints in different aspects, including learning, control and data modality. To begin with, a linear latent dynamics may allow to analyze and introduce a stability regulation criterion, with the potential of characterizing more robust goal-directed behaviors. Secondly, we plan to explore auxiliary tasks, as well as associated learned representation and constraints like III-B and III-C, to examine their impacts on the performance of addressing general motor control such as non-prehensile manipulation. Last but not the least, it would be interesting to investigate how rich data of other modalities, such as images or texts, can be leveraged by robots to reason about entity relations as well as the external world, so as to develop motor task skills from limited experience.

References

[1] S. Calinon. Robot learning with task-parameterized generative models. In *Proceedings of the International Symposium of Robotics Research (ISRR)*, 2015.

[2] S. Calinon, I. Sardellitti, and D. G. Caldwell. Learning-based control strategy for safe human-robot interaction exploiting task and robot redundancies. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 249–254, Oct 2010. doi: 10.1109/IROS.2010.5648931.

[3] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2017.

[4] Ioannis Havoutis and Subramanian Ramamoorthy. Motion planning and reactive control on learnt skill manifolds. *International Journal of Robotics Research*, 32 (9-10):1120–1150, 2013.

[5] Maximilian Karl, Maximilian Soelch, Justin Bayer, and Patrick van der Smagt. Deep variational bayes filters: Unsupervised learning of state space models from raw data. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2017.

[6] S. M. Khansari-Zadeh and A. Billard. Learning Stable Non-Linear Dynamical Systems with Gaussian Mixture Models. *Transactions on Robotics*, 2011.

[7] D P. Kingma and M. Welling. Stochastic gradient vb and the variational auto-encoder. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2014.

[8] J. R. Medina, D. Sieber, and S. Hirche. Risk-sensitive interaction control in uncertain manipulation tasks. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 502–507, May 2013.

[9] Marc. H. Raibert and John. J. Craig. Hybrid position/force control of manipulators. *ASME Journal of Dynamical Systems, Measurements and Control*, 103(2): 126–133, 1981. doi: 10.1115/1.3139652.

[10] Joel Rey, Klas Kronander, Farbod Farshidian, Jonas Buchli, and Aude Billard. Learning motions from demonstrations and rewards with time-invariant dynamical systems based policies. *Autonomous Robots*, 42(1): 45–64, Jan 2018.

[11] Aviv Tamar, Yi Wu, Garrett Thomas, Sergey Levine, and Pieter Abbeel. Value iteration networks. In *Proceedings of Neural Information Processing Systems (NIPS)*, pages 2154–2162. Curran Associates, Inc., 2016.

[12] Joshua Tobin, Rachel Fong, Alex K Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 23–30, 2017.

[13] M. Watter, J T. Springenberg, J. Boedecker, and M A. Riedmiller. Embed to control: A locally linear latent dynamics model for control from raw images. *CoRR*, abs/1506.07365, 2015.

[14] Markus Wulfmeier, Ingmar Posner, and Pieter Abbeel. Mutual alignment transfer learning. In *Proceedings of Conference on Robot Learning (CORL)*, 2017.

[15] H. Yin, A. Paiva, and A Billard. Learning cost function and trajectory for robotic writing motion. In *Proceedings of IEEE International Conference on Humanoid Robots (Humanoids)*, Madrid, Spain, 2014.

[16] H. Yin, F. S. Melo, A. Billard, and A. Paiva. Associate latent encodings in learning from demonstrations. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, San Francisco, USA, 2017.

[17] Brian D Ziebart, Andrew L Maas, J Andrew Bagnell, and Anind K Dey. Maximum entropy inverse reinforcement learning. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 1433–1438, 2008.